利用改进蚁群算法的重叠社团检测分析方法 *

许 英

(新疆财经大学 应用数学学院, 乌鲁木齐 830012)

摘 要：针对重叠社团检测准确率提升问题，提出了一种基于改进蚁群算法的新型重叠社团检测算法。该算法包含位置初始化、运动和后处理三个阶段，分别通过初始位置识别与标签列表存储、基于节点间相似度的启发式信息重定义、合作保持标签列表等方式，使算法在合成数据集与现实世界数据集中的重叠社团与节点检测方面具有更好的性能。实验结果表明，在合成网络与现实世界网络平台上使用不同检测算法，所提出的方法对重叠社团与重叠节点的检测准确率较传统检测方法来说更高，因而对重叠社区检测问题求解与理解网络功能结构具有重要的参考与借鉴意义。

关键词：重叠社团与节点检测；改进蚁群算法；启发式信息重定义；标签列表迭代更新

中图分类号：TP311.1 doi: 10.19734/j.issn.1001-3695.2018.10.0803

Novel algorithm of overlapping community detection and analysis with improved ant colony algorithm

Xu Ying

(*School of Applied Mathematics*, *Xinjiang University of Finance & Economics*, *Urumqi* 830012, *China*)

Abstract: This paper proposes a new detection algorithm for overlapping communities based on ant colony algorithm to improve the detection accuracy of overlapping communities. The algorithm consists of three stages: position initialization, motion and post-processing. The algorithm achieves better performance through initial position identification and tag list storage, heuristic information redefinition based on similarity between nodes, and cooperative tag list preservation in synthetic datasets and real-world datasets of overlapping communities and overlapping nodes detection. The detection performance of different detection algorithms on synthetic and real world network platforms shows that the proposed overlapping community detection and analysis method based on improved ant colony algorithm has good accuracy and analysis performance. The method can be used for reference in solving the current overlapping community detection problem and understanding the functional structure of the network.

Key words: overlapping community and node detection; improved ant colony algorithm; heuristic information redefinition; tag list iteratively update

0 引言

社团定义为网络结构中依据具备共享属性的顶点[1]。类似于社会网络中一个人具有不同角色，在标准化网络分区中，分区中每个顶点被分配给一个社团[2]，且被社团之间共享。此外，由于不同节点在网络结构和功能上具有不同作用，所以检测网络中的重叠社团可更深入地了解网络的功能与结构。因此，近年来重叠社团检测研究已引起学术界极大关注。众多具备良好分析性能的算法相继提出，并广泛应用于延迟容忍网络[3,4]、推荐系统[5]等领域。

当前社团检测算法大致可分为五类[6]，即派系过滤算法（clique percolation method，CPM）、局部扩展和优化算法、链接分区算法、模糊检测算法和基于代理的算法。文献[7]将索引局部邻接表示法引入社团检测时的个体表示中，将社团结构分析转换为整数优化问题，并据此提出基于差分进化的社团检测算法。文献[8]基于群体智能思想提出一种自组织的重叠社团结构分析算法，但智能体间收敛精度对检测性能影响较大。文献[9]从有效融合两类不同的异质信息研究出发，提出了一种基于交互行为和连接分析的社交网络社团检测方法，但检测性能很大程度依赖于数据属性。文献[10]提出一种局部优先的动态网络重叠社团演化分析方法，但其全局最优性有待验证。文献[11]提出了一种使用最优特征向量的谱二分社团检测方法，在上述重叠社团检测算法中，虽然可以一定程度上很好地解决重叠社区检测分析问题，但均忽略了迭代过程中节点标签变更对检测性能的影响。

因此，借鉴蚁群算法中信息素更新与运动路径选择策略，本文提出一种基于改进蚁群算法的重叠社区检测算法(ant colony based algorithm，AntCBO)。在建立基于蚁群算法的重叠社团检测框架基础上，通过基于节点间相似度的启发式信息重定义与基于标签列表更新存储的蚁群路径选择两个环节实现重叠社团检测，基于合成数据集和现实世界数据集平台开展检测算法性能分析。

1 AntCBO 重叠社区检测算法基本架构

网络抽象可建模为无向图 $G=(V, E)$，其中：V、E 分别为节点与边集合。设每个节点 v 都具有唯一标签（identification, ID）l_v，并记 $N(v)$为节点 v 的邻域。重叠社团检测目的为，寻求某种方法，将抽象得到的无向图 G 分割为一系列小集群($C_1, C_2, \ldots, C_m$)，使每个集群中具有相同的节点。数学描述为

收稿日期：2018-10-16；**修回日期：**2018-12-12 **基金项目：**国家自然科学基金资助项目（11461063）

作者简介：许英（1981-），女，新疆乌鲁木齐人，副教授，博士，主要研究方向为复杂网络理论及其应用(lu77320jiang68@163.com).

$$\cup_{1\le i\le m} C_i = G 且 \exists i,j\in m, C_i \cap C_j \ne \varnothing \quad (1)$$

1.1　算法框架

提出的 AntCBO 算法框架如算法 1 所示，包含四个模块，即参数初始化、蚂蚁位置初始化、运动路径决策与后处理。基本流程为：a）初始化参数，视每个节点为蚂蚁个体，并基于信息素和提出的启发式信息计算转移概率矩阵；b）初始化蚂蚁位置并确定对应的 ID 列表；c）所有蚂蚁基于转移概率模型移动位置，此过程中每个蚂蚁将节点标签从一个节点传递到另一个节点并更新信息素；d）满足终止条件时，保存蚂蚁个体当前标签列表，得到各节点标签序列；e）采用后处理机制将网络划分为重叠社区。

算法 1　AntCBO 算法框架

输入：复杂网络 *G*=(*V*, *E*)。

输出：一组重叠社团 C。

```
1 参数初始化：迭代次数最大值 T；蚂蚁数量 n；初始信息素值 τ；信息素挥发率 ρ；阈值 ϑ；信息素增加量 x；信息素阈值 b。
根据式（1）~式（3）计算转移概率矩阵。
2 蚂蚁位置初始化
3 蚂蚁移动：  for t=1 TO T DO
    for i=1 to n do
      每个蚂蚁 i 按照转移概率矩阵移动
      信息素扩散和更新
    end for
  end for
4 后处理
```

1.2　AntCBO 算法的改进

对 AntCBO 来说，$\tau_{ij}(t)$，$(i, j=1, 2,\dots, n$ 且 $i\ne j)$之间的相对大小会直接影响蚁群位置之间的转移概率，从而直接影响可行解的质量。这是由于搜索算法进行过程中，早起分散分布的信息素将逐渐集中至部分边上，从而不断强化搜索方向。当某些边的信息强度明显高出其他边时会导致可行解构造时选择的边过于接近导致解结构过于相似，从而算法陷入局部最优解。而避免算法陷入局部最优解的根本方法是增加可行解的多样性。其核心是使信息素在边上的分布不至过于集中，从而使得更多的边能够以较高的概率参与构造可行集过程。即既充分利用蚁群算法的正反馈机制加快搜索进程，又必须尽可能地过大算法可行解的搜索区域，使用更多边形成新的可行解。遵循这一思路，本文提出直接交换部分变上信息素的方法改变 $\tau_{ij}(t)$在不同边上的分布，从而避免算法后期陷入局部最优解。具体过程为：每个节点设定交换概率 p_i, $i=1, 2,\dots, n$，并生成随机数 $r_i \sim U(0, 1)$。若 $r_i\le p_i$，则在节点 i 与其他节点可以形成的 n-1 条边中，随机选择一定数量的边，两两互换对应的信息素值；若 $r_i>p_i$，则不作上述处理。

此外，信息素的挥发率也会影响信息素分布。传统蚁群算法中信息素挥发率 ρ 往往被设置为常数，且每条边的挥发率相同。此种操作会导致部分边用于构造可行解的概率过大，不利于其他边参与构造新可行解。因此，本文采用变参挥发率来避免最优路径上的某些边由于信息素强度过低而失去选择机会，其计算公式如下：

$$\rho_{ij}(t)=1 \qquad t\le t_{early} \quad (2)$$

$$\rho_{ij}(t)=\begin{cases} k_1 & \tau_{ij}(t)\le C \\ k_2 & else \end{cases} \qquad t>t_{early} \quad (3)$$

其中：$\rho_{ij}(t)$表示 t 时刻边 ij 上信息素挥发率；t_{early} 是早期搜索时间；$k_1, k_2\in(0,1)$且 $k_1>k_2$；此外 k_1、k_2 两者数值不宜差别过大，否则会导致交换后信息素挥发过快或过慢影响前述交换边效果；C 为介于 $\tau_{ij}(t)$均值与最大值之间的常数。

1.3　转移概率矩阵计算方法

由于直接模拟蚂蚁觅食行为进行社团相似节点聚类，算法核心为设计合理的启发信息与信息素更新策略计算转移概率矩阵，为蚁群移动路径提供决策依据。结合应用背景，首先采用任意节点 i 与 j 间相似性度值来更新启发信息，如式(4)5)所示。

$$comm(i,j)=neibour(i)\cap neibour(j) \quad (4)$$

$$\eta_{ij}=\gamma\times\frac{|comm(i,j)|}{|neibour(i)\cup neibour(j)|}+ (1-\gamma)\times\frac{2\times|E(comm(i,j))|}{(|comm(i,j)|)(|comm(i,j)-1|)} \quad (5)$$

其中：$|comm(i,j)|$ 表示节点 i 和 j 共同邻居的数量；$neibour(i)$ 表示节点 i 的邻居；$neibour(j)$ 表示节点 j 的邻居；$|E(comm(i,j))|$ 表示连接每个共同邻居节点的边的数量。从式(3)可知，启发式信息更新策略由两项组成，并通过权重系数 γ 控制两者在启发信息中的重要程度。其中，第一项为顶点 i 和 j 的公共邻居数量除以两个顶点邻居的并集数，它度量了两节点间平均公共邻居节点数；第二项是公共邻居之间的连边数量除以它们之间可能存在的连边数量，并且它量化了公共邻居彼此之间的连接距离。其实际物理意义为，若两邻居节点具有很多公共邻居且公共邻居连接距离较近，其属于同一社团概率更高。计及信息素与启发信息更新策略的状态转移概率矩阵如式(4)所示。

$$p_{i,j}=\begin{cases}\dfrac{\tau(i,j)^{\alpha}\eta(i,j)^{\beta}}{\sum_{q\in neibour(i)}\tau(i,j)^{\alpha}\eta(i,j)^{\beta}} & j\in neibour(i) \\ 0 & else\end{cases} \quad (6)$$

其中：α 和 β 表征信息素与启发信息在迭代过程中强化程度。

2　所提 AntCBO 算法步骤

2.1　蚂蚁位置初始化

此步骤首先进行蚂蚁（即节点）位置与对应的标签列表的初始化过程。即网络 G 按各节点的度[12]降序排列后计算各节点间的公共邻居并划分为初始社团 sg，根据初始社团各节点 ID 生成相应的标签列表，并将蚂蚁个体放置在初始集群的节点中。具体步骤如下所示。

算法 2　步骤一：蚂蚁位置初始化

输入：一个复杂网络 *G*(*V*, *E*)。

输出：蚂蚁的位置 *Cnode* 以及存储在每个节点 *v* 的标签列表 *L* (*v.L*)。

```
令 Cnode、CN、v.L 为空集
每个节点 v 具有一个 ID
∀v∈V : v.available = true
根据节点的度对所有节点进行降序排序得到列表 L
for each vᵢ∈L DO
sg ← ∅
if  vᵢ.available 且 d(vᵢ) ≥ 3  then
找到 vᵢ 的邻居中度最大对的节点 vⱼ 并令 vⱼ.availabe ← true
sg ← sg ∪ {vᵢ, vⱼ}
计算 vᵢ 和 vⱼ 的共同邻居 comNei
while  comNei ≠ ∅  do
for each  vₘ  in  comNei  do
sg ← sg ∪ {vₘ}
Z = neibour(vₘ) ∩ comNei 并用 Z 代替 comNei
end for
end while
end if
CN ← CN ∪ sg
```

```
for each v_x in sg do
v_x.l ← v_j.ID
end for
end for
Cnode 是 CN 中节点，不在 Cnode 中的节点使用其本身 ID 初始化
```

2.2　蚂蚁的移动

每个蚂蚁都会根据转移概率将 ID 在两节点间转移。为提高蚂蚁遍历整个网络抽象图 G 的几率，中蚂蚁移动规则设为：如果随机数 r 小于 0.1，则蚂蚁将选择具有最大概率的邻居作为下一个要访问的节点；否则蚂蚁随机选择除具有最大转移概率的邻居节点作为访问节点。而蚂蚁需访问多个具有相同转移概率的 ID 时，将随机选择其一。

迭代结束时，蚂蚁从当前节点移动至下一节点。此外，每个蚂蚁从当前节点的标签列表中选择其一存放至下一节点标签列表中。信息素则存放与蚂蚁在不同节点转移的路径之上。信息素挥发过程则在每次迭代过程结束之时。蚂蚁运动策略如算法 3 所示。

算法 3　步骤二：蚂蚁的移动

```
while 不满足终止条件 do
for 每只蚂蚁 DO
蚂蚁基于转移概率矩阵从节点 i 移动到节点 j
蚂蚁从一个节点移动到下一节点，并从当前节点到下一个节点获得一个标签，将这一标签保留在下一节点标签列表
该边的信息素量增加 x
if 信息素值超过阈值 b then
将信息素的值设置为 b
end if
所有的蚂蚁移动后，所有边的信息素按 ρ 挥发，  τ_new = (1−ρ)×τ
end for
重新计算转移概率矩阵
end while
```

此外，为避免算法过早收敛到非全局最优解，规定：当信息素超出预定值 b 时，应将边缘信息素限制到最大值。通过这种方式，可以有效地避免一条路径上的信息量远远大于其他路径上的信息量，并且可消除所有蚂蚁集中在同一条路径的情况。

2.3　后处理

迭代终止后，统计各节点中标签列表各从属值出现的频率。若某一值出现频率在整个标签列表中取值较大，则该节点属于该标签的概率更高。为提高重叠社团检测准确率，可采用算法对待检测网络进行多次扫描，以确定最终的社团划分。后处理步骤如算法 4 所示。其中：$|N^{l(v)}|$ 表示具有标签 l 的节点 v 的邻居节点的数量；sum 表示节点 v 的邻居节点中的所有可能的标签的数量。

算法 4　步骤三：后处理

输入：节点及其标签列表。

输出：重叠社团的集合 C。

```
首先根据标签列表中频率从最大的标签获得每个节点的标签
for each 节点 v do
找到节点 v 邻居节点 N(v) 并记录其邻居的标签集合 L
for each 标签 l in L do
if |N^l(v)|/sum > θ then
l 是节点 v 所从属的一个社团
end if
end for
end for
删除所有包含在大社团中的小社团
返回重叠社团集合 C
```

3　实验结果与分析

为验证 AntCBO 算法在求解重叠社区检测问题的准确度和分析性能，分别在合成网络和现实世界网络环境下进行对比实验实验，实验中以派系过滤算法（cluster percolation method，CPM）和社区重叠传播算法（community overlap propagation algorithm，COPRA）作为 AntCBO 的对比算法，通过实验对比算法在重叠社区检测准确度的优劣性。

对上述三种算法进行实验时，需要对算法参数进行设置，实验参数设置如下：CPM 算法中，对于合成网络数据集，k=3，对于真实世界数据集，k 从 3 到 10 变化；COPRA 算法中，对于合成网络数据集，v=8，对于真实世界数据集，v 在 1 到 9 之间变化；而 AntCBO 的实验参数如表 1 所示。

表 1　AntCBO 算法的参数

Table 1　Parameters of antcbo algorithm

描述	参数表示	值
最大迭代次数	T	500
初始化信息素值	τ	8
信息素衰减	ρ	0.12
阈值	θ	(0.08, 0.5)
信息素增加	x	0.15
信息素阈值	b	20

3.1　合成网络实验

合成网络实验通过广泛采用局部最大化（local fitness raximization, LFR）基准来生成，可更好地实现节点数量和节点满足幂律分布，因此 LFR 更接近真实的社交网络。在实验中，本文通过 LFR 基准生成四个数据集，这些数据集的参数如表 2 所示。

表 2　合成网络的参数

Table 2　Parameters of synthetic network

数据集	节点数量	社团尺寸范围	μ	O_n	O_m
1	1500	10~50	0.08	100	2~6
2	1500	10~50	0.32	100	2~6
3	4500	10~50	0.08	500	2~6
4	4500	10~50	0.32	500	2~6

3.2　重叠社团识别准确度分析

合成网络实验中，采用扩展归一化互信息（normalized mutual information，NMI）来对比三种方法对重叠社团识别的准确性。NMI 可以表征检测到的分区和真实分区之间的相似性，其值在 0~1 之间变化，且 NMI 值越高，分区结果越好。三个合成数据集上的三种算法的最佳 NMI 值如图 1 所示。对于分区 C' 和 C''，NMI 值由式(7)和(8)确定。

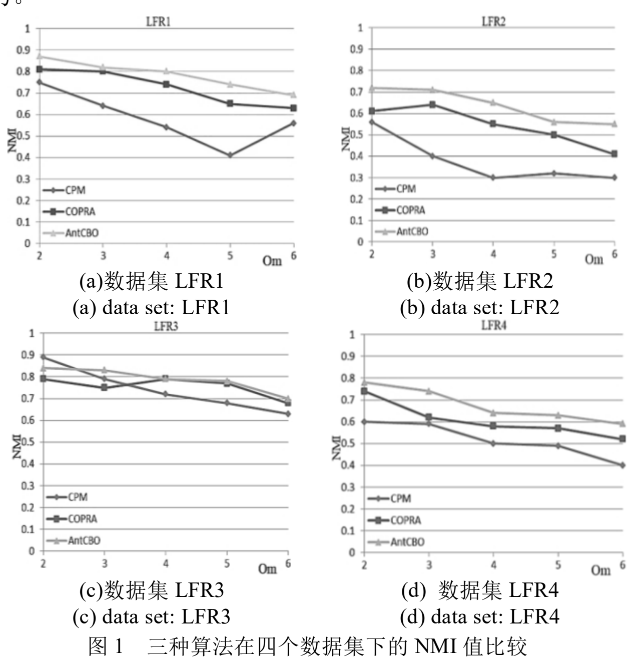
$$NMI = 1 - \frac{1}{2}[H(X|Y) + H(Y|X)] \tag{7}$$

$$H(X|Y) = \frac{1}{|C'|}\sum_k \frac{H(X_k|Y)}{H(X_k)} \tag{8}$$

其中：X、Y 表示分区 C' 和 C'' 相关联的随机变量；$H(X|Y)$是节点属于 Y 的前提下还属于 X 的归一化条件熵，也可以以同样的方式定义为节点属于 X 的前提下还属于 Y 的归一化条件熵。

为验证算法的普适性，首先对比分析同一数据集下不同算法的重叠社团检测性能。从图 1(a)~(d)可以看出，对于不同的网络规模和不同的社区范围大小，成员数量 O_m 从 2~6 变化时，本文所提 AntCBO 算法在 LFR1、LFR2 和 LFR4 数据集中的 NMI 值始终高于其他两种算法，而对于 LFR3 数据集而言，虽然在成员数量较小时 NMI 值低于 CPM 算法，但

成员数量增大后其 NMI 值同样高于另外两种算法。这说明本文所提 AntCBO 算法对于成员数量较高时具有很强的检测性能，而低成员数量时虽性能有所下降但仍具有较高的辨识能力。



(a)数据集 LFR1　(a) data set: LFR1

(b)数据集 LFR2　(b) data set: LFR2

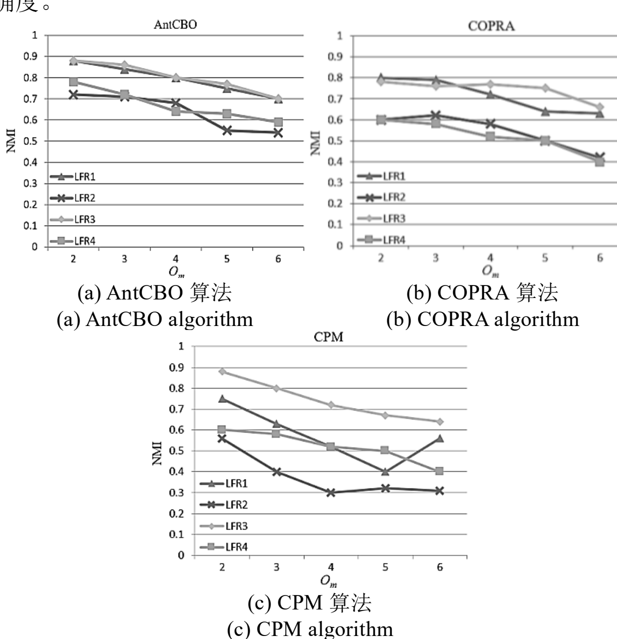
(c)数据集 LFR3　(c) data set: LFR3

(d) 数据集 LFR4　(d) data set: LFR4

图 1　三种算法在四个数据集下的 NMI 值比较

Fig. 1　Comparison of NMI values of three algorithms under 6 data sets

为验证算法的健壮性并分析网络规模对重叠社团检测性能的影响，对比同一算法随网络规模增大时的检测性能。图 2(a)~c)示出了三种算法在数据集 LFR1~LFR4 下的 NMI 值变化情况。可以看出，随着成员数量的增多和网络规模的扩大，上述三种算法的 NMI 值均出现了不同程度的下降。从下降趋势来看，虽然 AntCBO 算法与 COPRA 算法的下降幅度相近（均在 0.2 左右），但本文所提 AntCBO 算法在网络规模较大(即 LFR4，O_m=6)时其 NMI 指标仍比 COPRA 算法核 CPM 算法分别高 30.9%和 83.3%，从而说明本文所提 AntCBO 算法对于网络规模大、成员数量多的重叠社团检测具有更高的准确度。



(a) AntCBO 算法　(a) AntCBO algorithm

(b) COPRA 算法　(b) COPRA algorithm

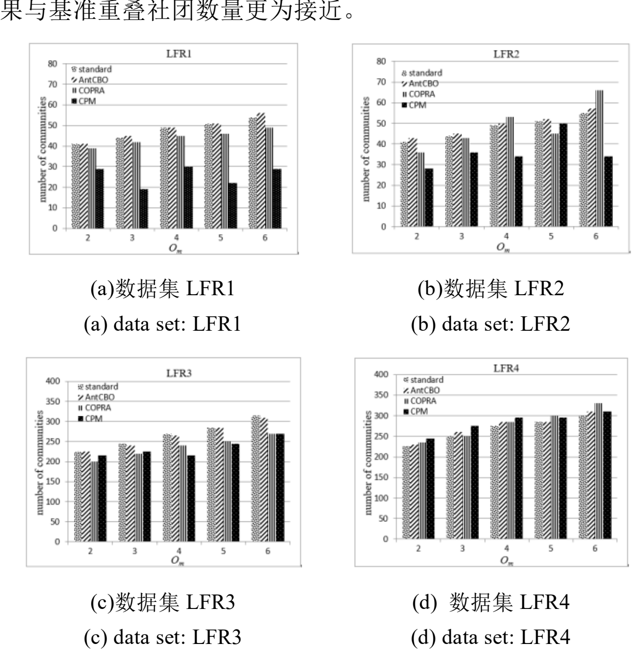
(c) CPM 算法　(c) CPM algorithm

图 2　三种算法在 LFR1~LFR4 上的 NMI 值

Fig. 2　NMI values of three algorithms on LFR1~LFR4

3.3　重叠社团检测数量分析

进一步分析本文所提 AntCBO 算法对重叠社团检出数量的性能。图 3 示出了不同算法下，检测出重叠社团数量与基准测试中标准社团数量的对比示意图。图 3(a)~(d)可以看出，虽然三种算法在网络规模较小且成员数量较少时均能较好地反映出重叠社团数量的真实值；但随着网络规模和成员数量增大时，相较于另外两种算法，本文所提 AntCBO 算法的结果与基准重叠社团数量更为接近。



(a)数据集 LFR1　(a) data set: LFR1

(b)数据集 LFR2　(b) data set: LFR2

(c)数据集 LFR3　(c) data set: LFR3

(d) 数据集 LFR4　(d) data set: LFR4

图 3　在四个数据集上通过三种算法检测到的社区数量

Fig. 3　Number of communities detected by 3 algorithms on four data sets

3.4　重叠社团检测准确度分析

实验中为了方便地分析算法对网络重叠社团的识别准确度，定义函数 *Fscore* 用于测量检测重叠社团算法的准确性，它是准确率（*precision*）和召回率（*recall*）的调和平均值，其定义公式如式(9)所示。

$$Fscore=\frac{2*precision*recall}{precision+recall} \tag{9}$$

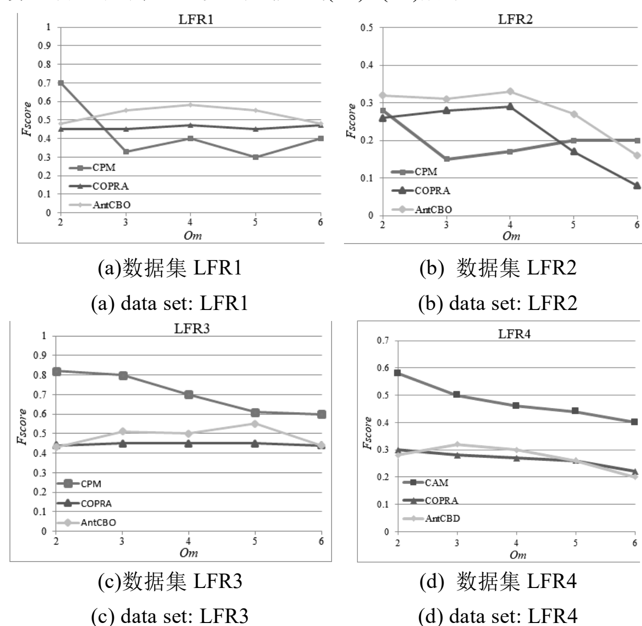
其中：*precision* 是正确检测到的重叠节点的数量除以检测到的重叠节点的所有数量；*recall* 被定义为正确检测到的重叠节点的数量除以重叠节点的真实数量所得到的值。

图 4 示出了同一数据集下不同检测算法对重叠社团的辨识准确度。从图中可以看出，本文所提 AntCBO 算法在 LFR1 和 LFR2 的检测准确度较好（即 *Fscore* 函数值更高），故在网络规模较小时，本文所提 AntCBO 算法能够较准确地辨识出重叠社团；而当网络规模较大时（如 LFR3 和 LFR4），CPM 算法的 *Fscore* 函数值相比于 AntCBO 算法和 COPRA 算法更高，说明其对重叠社团的检测更为准确。此外，从图 4(a)~(d) 可以看出，本文所提 AntCBO 算法比 COPRA 的检测准确度均要好。结合 3.1.1 和 3.1.2 节分析，虽然 CPM 算法在网络规模较大和成员数量较多时的重叠社团检测准确率高，但本文所提 AntCBO 算法在重叠社团检测数量上更接近真实网络。故总体而言，本文所提 AntCBO 算法性能强于另外两种算法。

3.5　真实世界社交网络

本节讨论在真实世界社交网络中所提出的 AntCBO 算法对重叠社团的检测性能。相关参数如表 3 所示。由于真实世界网络的重叠社区结构未知，所以 NMI 值不能用于测量算法的性能，实验中采用 Q_{ov} 来评估算法性能。Q_{ov} 的定义为重叠

社区检测的经典模块性的扩展，它考虑了每个顶点所属的社区数量以及每个社区的成员资格程度，Q_{ov} 的值越高，说明算法分区越好。主要公式如式(10)~(15)所示。



(a)数据集 LFR1　　(b) 数据集 LFR2

(a) data set: LFR1　　(b) data set: LFR2

(c)数据集 LFR3　　(d) 数据集 LFR4

(c) data set: LFR3　　(d) data set: LFR4

图 4　基于 LFR1~LFR4 的重叠节点检测评估

Fig. 4　Evaluation of overlapped node detection and based on LFR1~LFR4

表 3　真实世界社交网络参数

Table 3　Digest ofreal world social networking

数据集	节点数量	边数量	平均度	社团数量
Karate[13]	33	78	4.6	2
Dolphins[14]	60	160	5.2	2
Football[15]	116	615	10.8	12

$$f(x)=60x-30 \tag{10}$$

$$F(\alpha_{i,c},\alpha_{j,c})=\frac{1}{(1+e^{f(\alpha_{i,c})})(1+e^{-f(\alpha_{j,c})})} \tag{11}$$

$$\beta_{l(i,j),c}=F(\alpha_{i,c},\alpha_{j,c}) \tag{12}$$

$$\beta^{out}_{l(i,j),c}=\frac{\sum_{j\in V}F(\alpha_{i,c},\alpha_{j,c})}{|V|} \tag{13}$$

$$\beta^{in}_{l(i,j),c}=\frac{\sum_{i\in V}F(\alpha_{i,c},\alpha_{j,c})}{|V|} \tag{14}$$

$$Q_{ov}=\frac{1}{m}\sum_{c}\sum_{i,j\in V}\left[\beta_{l(i,j),c}A_{ij}-\frac{\beta^{out}_{l(i,j),c}k^{i}_{out}\,\beta^{in}_{l(i,j),c}k^{j}_{in}}{m}\right] \tag{15}$$

根据实验结果，Q_{ov} 值的对比数据如表 4 所示。由表 4 分析可得，AntCBO 算法在 karate、dolphins、football 网络上的表现其优于 COPRA 和 CPM。

表 4　真实世界网络中不同算法的 Q_{ov} 值

Table 4　Values of Q_{ov} with different algorithms in real-world networks

数据集	AntCBO	COPRA	CPM
Karate	0.73	0.52	0.53
Dolphins	0.73	0.70	0.67
Football	0.70	0.68	0.63

综上所述，对于大多数真实世界网络数据集，AntCBO 算法可以获得比 COPRA 和 CPM 更高的 Q_{ov} 值，这也说明 AntCBO 算法相比于另两种算法具备更好的性能。

4　结束语

通过现有对重叠社区检测算法的研究，本文提出了一种基于蚁群的重叠群体检测算法。该算法包括参数初始化、蚂蚁的位置初始化、蚂蚁的移动和后期处理等阶段和步骤。以此同时，通过在合成和现实世界数据集上进行的实验分析，对算法所能实现的性能作了详细验证。实验结果表明，相对于 CPM 和 COPRA，所提出的基于蚁群的重叠群体检测算法在社区检测质量方面具有良好的性能，对当前重叠社区检测问题的求解和网络中功能结构的理解具有重要的参考和借鉴意义。

参考文献：

[1] Wang Xiaofeng, Liu Gongshen, Li Jianhua. Overlapping community detection based on structural centrality in complex networks [J]. IEEE Access, 2017, 10 (5): 25258-25269.

[2] Lyzinski V, Tang M, Athreya A, *et al.* Community detection and classification in hierarchical stochastic blockmodels [J]. IEEE Trans on Network Science & Engineering, 2017, 4 (1): 13-26.

[3] 黄宏程，熊忠阳，胡敏，等. 节点移动状态感知的社会化延迟容忍网络路由策略 [J]. 计算机应用研究, 2017, 34 (6): 1825-1829. (Huang Hongcheng, Xiong Zhongyang, Hu Min, *et al.* Routing strategy based on change perception of node mobility characteristic in DTN [J]. Application Research of Computers, 2017, 34 (6): 1825-1829.)

[4] 侯林清，蔡英，范艳芳，等. 移动社交网中基于兴趣社区的消息传输方案 [J]. 计算机科学, 2018, 45 (6): 105-110. (Hou Linqing, Cai Ying, Fan Yanfang, *et al.* Interest community based message transmission scheme in mobile social networks [J]. Computer Science, 2018, 45 (6): 105-110.)

[5] 陈东明，严燕斌，黄新宇，等. 基于二分网络社团划分的推荐算法 [J]. 东北大学学报：自然科学版 , 2018, 39 (8): 1103-1107. (Chen Dongming, Yan Yanbing, Huang Xinyu, *et al.* Recommendation algorithm based on community detection in bipartite networks [J]. Journal of Northeastern University :Natural Science , 2018, 39 (8): 1103-1107.)

[6] Xie Jierui, Kelley S, Szymanski B K. Overlapping community detection in networks: the state of the art and comparative study [J]. ACM Computing Surveys, 2011, 45 (4): 1-35.

[7] 孙韩林，马素刚，王忠民. 一种基于差分进化的社团检测算法 [J]. 软件工程, 2018, 21 (1): 1-6. (Sun Hanlin, Ma Sugang, Wang Zhongmin, A community detection algorithm using differential evolution [J]. Software Engineering, 2018, 21 (1): 1-6.)

[8] 孙韩林，马素刚，王忠民. 一种基于群体智能的自组织重叠社团结构分析算法 [J]. 计算机应用研究, 2019, 36 (5) , doi: 10. 3969//j. issn. 1001-3695. 2017. 11. 0733. (Sun Hanlin, Ma Sugang, Wang Zhongmin. Self-organizing overlapping community structure analysis algorithm based on swarm intelligence [J]. Application Research of Computers, 2019, 36 (5) , doi: 10. 3969//j. issn. 1001-3695. 2017. 11. 0733.)

[9] 李鹏，李英乐，王凯，等. 基于交互行为和连接分析的社交网络社团检测 [J]. 计算机科学, 2017, 44 (7): 197-202. (Li Peng, Li Yingle, Wang Kai, *et al.* Community detection based on user interaction and link analysis in social networks [J]. Computer Science, 2017, 44 (7): 197-202.)

[10] 彭焱，溪利亚. 局部优先的动态网络重叠社团及其演变模式检测 [J]. 计算机工程, 2016, 42 (12): 188-195. (Peng Yanm Xi Liya.

Local-first detection of overlapping community and its evolution pattern in dynamic networks [J]. Computer Engineering, 2016, 42 (12): 188-195.)

[11] 周旸，陈晓云，程建军，等. 基于最优特征向量的谱二分社团检测方法 [J]. 计算机科学与探索，2017, 11 (12): 1897-1906. (Zhou Yang, Chen Xiaoyun, Cheng Jianjun, *et al.* Bisection spectral community-detection methods using optimal eigenvectors [J]. Journal of Frontiers of Computer Science & Technology, 2017, 11 (12): 1897-1906.)

[12] 邓承刚，李国徽，杨浩，等. 基于指引因子蚂蚁算法的仿真资源分配策略 [J]. 华中科技大学学报 :自然科学版 , 2015, 1 (9): 1-6. (Deng Chenggang, Li Guohui, Yang Hao, *et al.* Simulation resource allocation strategy based on guidance factor ant algorithms [J].Journal of Huazhong University of Science and Technology :Nature Science Edition , 2015, 1 (9): 1-6.)

[13] 陆亿红，张振宁，杨雄. 一种基于节点特征向量的复杂网络社团发现算法 [J]. 计算机科学，2017, 44 (6): 419-423. (Lu Yihong, Zhang Zhenning, Yang Xiong. Community structure detection algorithm based on nodes'eigenvectors [J]. Computer Science, 2017, 44 (6): 419-423.)

[14] Allen S J, Pollock K H, Bouchet P J, *et al.* Preliminary estimates of the abundance and fidelity of dolphins associating with a demersal trawl fishery [J]. Scientific Reports, 2017, 7 (1): 4995-5002.

[15] 王玉英，何汶坤，史加荣. 基于社团密度的社团发现算法 [J]. 计算机应用研究，2017, 34 (7): 1975-197. (Wang Yuying, He Wenkun, Shi Jiarong. Community detection algorithm based on community density [J]. Application Research of Computers, 2017, 34 (7): 1975-197.)